

### **REMARKS**

Claims 1-40, 42 and 43 are currently pending in the subject application and are presently under consideration. Claims 1, 13, 29, 42, and 43 have been amended as shown on pages 2-10 of the Reply.

Applicant's representative thanks Examiner Hicks for the courtesies extended during the telephonic interview conducted on April 1, 2008. During the discussion, the Examiner made a number of recommendations for claim amendments to overcome the rejections under 35 U.S.C. §101. The Examiner's suggestions have been incorporated into the claim set herein. The participants also discussed the amendments proposed to more clearly distinguish the present claims from the cited references. Specifically, the feature of automatically populating a non-relevant training data set with unselected query results that are ranked higher than a selected query result was discussed in light of the search refinement system disclosed in Pazzani, *et al.* The Examiner, while indicating that the feature being disclosed appears to overcome the cited reference, asked that the wording of this feature in the independent claims be further modified to preclude interpretations broad enough to encompass Pazzani, *et al.* Accordingly, the claims herein have been further amended from their originally proposed form in view of the Examiner's indicated interpretation.

Favorable reconsideration of the subject patent application is respectfully requested in view of the comments and amendments herein.

#### **I. Rejection of Claims 1-21 and 43 Under 35 U.S.C. §101**

Claims 1-21 and 43 stand rejected under 35 U.S.C. §101 because the Examiner contends the claimed invention is directed to non-statutory subject matter. Independent claims 1, 13, and 43 have been amended to address the Examiner's concerns with regard to this rejection. It is therefore respectfully requested that this rejection be withdrawn.

#### **II. Rejection of Claims 1-40 and 42 Under 35 U.S.C. §103(a)**

Claims 1-40 and 42 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Lawrence in view of Pazzani, *et al.* It is respectfully submitted that this rejection should be withdrawn for at least the following reasons. Lawrence and Pazzani, *et al.*, individually or in combination, do not teach or suggest each and every feature set forth in the subject claims.

A factfinder should be aware, of course, of the distortion caused by hindsight bias and must be cautious of arguments reliant upon *ex post* reasoning. See *KSR v. Teleflex*, 550 U.S. \_\_\_, 127 S. Ct. 1727 (2007) citing *Graham v. John Deere Co. of Kansas City*, 383 U. S. 1, 36 (warning against a “temptation to read into the prior art the teachings of the invention in issue” and instructing courts to “guard against slipping into the use of hindsight” (quoting *Monroe Auto Equipment Co. v. Heckethorn Mfg. & Supply Co.*, 332 F. 2d 406, 412 (CA6 1964))).

The subject claims relate to a system and method for refining search query results from a general-purpose search engine based in part on the entry point through which the general-purpose search engine was accessed. When the search engine is accessed *via* an entry point and a search query is executed, the search query results obtained by the search engine can be passed to a tuning component associated with the entry point. The tuning component can filter and rank the search query results according to a statistical analysis that utilizes two distinct sets of data associated with the entry point and the context of the search: a first set of data expressly defined as relevant to the search context and a second data set expressly defined as non-relevant to the search context. Additionally, the first and second data set can be automatically trained based on observance of user selections. Specifically, when a user selects one of the filtered search query results presented by the tuning component, that result can be automatically recorded by the training component as relevant to the search context and added to the first data set, while results that had been ranked higher than the selected result can be automatically added to the category of non-relevant data. In particular, amended independent claim 1 recites, *user selection of a query result from a ranked list of the query results causes the selected result to be added to the first set of data and causes the results not selected by the user but ranked higher than the selected result to be automatically added to the second set of data.*

As conceded in the Office Action, Lawrence does not disclose these aspects of the subject claims. The Examiner contends that Pazzani, *et al.* makes up this deficiency. Pazzani, *et al.* presents an overview of an intelligent agent called Syskill & Webert used to develop and refine user profiles that infer websites of interest to the associated user. These user profiles can be revised and updated in response to feedback from the associated user regarding which websites are of interest and which are not, and these updated profiles can be used to predict

which websites will be of most interest to the user. Although Pazzini, *et al.* teaches that the algorithms used to update the user profiles employ a set of positive examples (*e.g.* websites the user is interested in) and negative examples (*e.g.* websites the user is not interested in), the cited reference indicates that these examples must be *explicitly selected* by the user in *both cases*. Section 2.1, paragraph 3, for example, explains that pages are rated by a user as being “hot” (*i.e.* interesting) or “cold” (*i.e.* not interesting), and these ratings are used to train the algorithm that refines the user profiles. Hence, websites that a user finds uninteresting for a given search session must be *explicitly selected* by the user for inclusion in the negative examples. In the Office Action, the Examiner appears to equate a website that is “rated as uninteresting” with “unselected.” However, this is not a true equivalence in the context of the cited reference, which teaches that uninteresting websites must be directly rated as such by the user in order to be included in the cited negative examples, and as such are not “unselected.” This is also underscored in paragraph 1 of the Introduction, which states that “an initial profile...is then revised when the user rates *visited* sites.” Clearly, Pazzini, *et al.* expects a user to visit (*i.e.* select) a website before rating the site as interesting or not interesting. The subject claims, by contrast, disclose that query results that are ranked higher than a selected result can automatically be included in the set of non-relevant training data without the need to visit or rank the non-relevant sites, thereby removing the burden of explicitly indicating uninteresting websites from the user.

As already noted, the subject claims disclose that the aforementioned first and second data sets can be associated with an entry point through which a general-purpose search engine is accessed. To this end, amended independent claim 1 further recites, *a tuning component that receives search query results of the general-purpose search engine and filters the search results based at least on criteria associated with the entry point through which the general-purpose search engine was accessed, the criteria comprises at least a first set of data categorized as relevant to a user's context and a second set of data categorized as non-relevant to the user's context*. As conceded in the Office Action, Lawrence does not teach or suggest the first and second data sets recited in the subject claims, and therefore does not disclose that such data sets can be *associated with an entry point* used to access the search engine. Contrary to the Examiner's assertions, Pazzini, *et al.* is also silent regarding such an association. Rather, the

cited reference teaches that the training data is associated with a *user profile*, which is not equivalent to an *entry point* through which a search engine is accessed.

Similarly, amended independent claim 13 recites, *a filter component that receives search query results of a general-purpose search engine and parses relevant and non-relevant results based on training data associated with the entry point that provides a link employed to traverse to the general-purpose search engine...and a ranking component that sorts the filtered results in accordance with the training data for presentation to a user, wherein a user clicking a link associated with a search result from the sorted results causes the result to be added to the first set of data and causes the results whose links were not clicked by the user but that are ranked higher than the clicked result to be automatically added to the second set of data.* As discussed *supra*, neither Lawrence nor Pazzini, *et al.* disclose these aspects of the subject claims.

Likewise, independent claim 22 recites, *automatically storing a first query result selected by a user in a first data set categorized as relevant; automatically storing at least one non-selected query result that is ranked higher than the first query result in a second data set categorized as non-relevant upon selection of the first query result; and including the first data set and second data set in the set of training data associated with the entry point employed to access the general purpose search engine.* Lawrence and Pazzini, *et al.* are silent regarding this method of building training data sets as already noted. The cited references also fail to disclose associating such training data with an entry point employed to access a general-purpose search engine, as also discussed above.

Amended independent claim 29 discloses similar features, reciting, *executing a search query with the general purpose search engine to obtain a ranked list of query results; selecting a link associated with a query result from the list; automatically adding the selected query result to the first set of data; and automatically adding non-selected results from the list that are ranked higher than the selected query result to the second set of data upon selection of the selected query result.* As already discussed these aspects are not disclosed by the cited references.

Moreover, amended independent claim 34 recites, *recording a first query result from a ranked list of query results returned from the executed query as relevant when a user views the document associated with the first query result; recording at least one second query result whose associated document was not viewed by the user but that is ranked higher than the first query result as non-relevant when the first result is selected for viewing by the user*, and as already noted, the cited references do not disclose such a method for training data sets.

Furthermore, amended independent claim 42 recites, *a component that ranks the filtered general-purpose search engine results according to the similarity of the search engine results to the training data sets, wherein selecting a link associated with a first search result from the ranked results causes the first result to be added to the first set of data and causes results that are ranked higher than the first result and have not been selected by the user to be automatically added to the second set of data*. The cited references are silent regarding these aspects, as noted above.

With further regard to the statistical analysis used to determine the relevance or non-relevance of a search result in view of the training data sets, the subject claims disclose that a probability distribution can be generated for each of the relevant data set, the non-relevant data set, and a returned result. A threshold can be defined between the relevant and non-relevant distributions, and the distribution for the returned result can be compared with this threshold to determine the relevance or non-relevance of the result. Additionally, the threshold can be biased to favor either the relevant or non-relevant distribution, according to preference. In particular, claim 25 recites, *the threshold employed to bias the decision to mitigate one of a result being deemed non-relevant when the result is relevant and a result being deemed relevant when the result is non-relevant*. While conceding that Lawrence fails to disclose such a threshold bias, the Examiner contends that Pazzani, *et al.* remedies this deficiency, arguing that the statistical methods employed by that cited reference must include a threshold to indicate the separation between relevance and non-relevance. However, Pazzini, *et al.* does not explicitly disclose *biasing* such a threshold to favor relevance or non-relevance, as recited in claim 25.

In view of at least the foregoing, it is respectfully submitted that Lawrence and Pazzini, *et al.*, individually or in combination, do not teach or suggest all aspects of independent claims 1, 13, 22, 29, 34, and 42 (and all claims depending there from), and as such fail to make obvious the present invention. It is therefore requested that this rejection be withdrawn.

### III. **Rejection of Claim 43 Under 35 U.S.C. §103(a)**

Claim 43 stands rejected under 35 U.S.C. §103(a) as being unpatentable over Lawrence in view of Taguchi, *et al.* as above and in further view of Hansen, *et al.* However, amended independent claim 43 recites, *means for ranking the filtered general-purpose search engine results based on a relevance of the general-purpose search engine results to the search context of the group of users and the entry point as determined by a comparison of the search engine results with the first and second sets of training data, wherein a user viewing a document associated with a first search result from the ranked results causes the first result to be added to the first set of data and causes the results that are unviewed but ranked higher than the first result to be automatically added to the second set of data.* As discussed in the previous section of the Reply, and as conceded in the Office Action, Lawrence does not teach or suggest this feature of training data sets. Taguchi, *et al.*, which relates to a file searching system that can retrieve search results from a document index and perform subsequent filtering on the retrieved results, also fails to disclose these aspects, since the filtering disclosed in that reference is based on a *user's access rights* to a retrieved document, and not on training data sets indicating relevant and non-relevant context data.

Hansen, *et al.* is likewise silent regarding the recording of higher ranked but unselected query results as non-relevant data. Hansen, *et al.* teaches the use of a PageRank algorithm that groups relevant queries together based on the similarities of respective search sessions. However, the cited reference does not contemplate *storing higher ranked but unselected search results* to any kind of data set, much less a specifically designated set of non-relevant data as disclosed in the subject claims. Indeed, Hansen, *et al.* explicitly discloses that only those search results that are selected by the user are considered during query clustering (see page 137, column 1, paragraph 4). Hence, Hansen expressly teaches away from the training techniques disclosed in amended independent claim 43.

In view of at least the foregoing, it is respectfully submitted that Lawrence, Taguchi, *et al.*, and Hansen, *et al.*, individually or in combination, do not teach or suggest each and every feature set forth in amended independent claim 43, and as such fails to anticipate the subject claim. Consequently, this rejection should be withdrawn.

**CONCLUSION**

The present application is believed to be in condition for allowance in view of the above comments and amendments. A prompt action to such end is earnestly solicited.

In the event any fees are due in connection with this document, the Commissioner is authorized to charge those fees to Deposit Account No. 50-1063 [MSFTP444US].

Should the Examiner believe a telephone interview would be helpful to expedite favorable prosecution, the Examiner is invited to contact applicant's undersigned representative at the telephone number below.

Respectfully submitted,

AMIN, TUROCY & CALVIN, LLP

/Himanshu S. Amin/

Himanshu S. Amin

Reg. No. 40,894

AMIN, TUROCY & CALVIN, LLP  
24<sup>TH</sup> Floor, National City Center  
1900 E. 9<sup>TH</sup> Street  
Cleveland, Ohio 44114  
Telephone (216) 696-8730  
Facsimile (216) 696-8731